



## Analysis and Classification of Arabic Newspapers' Facebook Pages using Text Mining Techniques

Said A. Salloum<sup>1,2</sup>, Chaker Mhamdi<sup>3,4</sup>, Mostafa Al-Emran<sup>4,5</sup>, Khaled Shaalan<sup>1</sup>  
ssalloum@uof.ac.ae; shaker@buc.edu.om; malemran@buc.edu.om; khaled.shaalan@buid.ac.ae

<sup>1</sup> The British University in Dubai, UAE.

<sup>2</sup> University of Fujairah, UAE.

<sup>3</sup> University of Manouba, Tunisia.

<sup>4</sup> Al Buraimi University College, Oman.

<sup>5</sup> Universiti Malaysia Pahang, Malaysia.

**Abstract.** Text mining and sentiment analysis have been the focus of research due to the increasing availability of opinion data on social networking websites. Current researches dealing with sentiment analysis aim at investigating people's behaviors. Social networks are rich sources of people's posts and comments regarding various topics which offer a good soil for research. Arabic language is the fifth among the most extensively utilized languages on the Internet. However, investigating social media posts and comments in Arabic language seems to be under-researched. Accordingly, this paper analyzes and classifies posts and comments of Arabic newspapers' Facebook pages. A total number of 24 Arab Gulf newspapers' Facebook pages were investigated where 62327 posts and 9372 comments in Arabic were studied and analyzed. Different text mining techniques were employed to analyze the extracted data. As far as Arab Gulf region is concerned, findings indicated that the UAE is the country that most frequently shares posts on Facebook, followed by Oman and KSA, respectively. Additionally, the findings reveal that videos are the most attracting post type on Arabic newspapers Facebook pages.

**Keywords:** Text mining, Sentiment Analysis, Newspapers, Facebook, Classification.

### 1. Introduction

Nowadays, due to the rapid pace of globalization and ever-increasing advances of information technologies which have obviously influenced various aspects of human lives, the environment of media practices has seen radical transformations. Today, news is disseminated freely and widely available to internet users with much more ease than it used to be with traditional media. Social media has evidently been changing the traditional pattern of media practices. As argued by Mhamdi (2016), "news media is increasingly a forum for information and debate with a non-linear flow of information and open-sourced journalism" (p. 272). This impact of social media in journalism practices is mainly attributed to its interactive feature and considerable role in communication and breaking news (Mhamdi, 2017). In addition, Al-Qaysi and Al-Emran (2017) pointed out that social media makes it easy to communicate with colleagues and relatives on anytime anywhere settings.

Facebook is the biggest social network with more than one billion clients. Facebook users spend approximately 120 minutes consistently to communicate with family and companions. The usage of some Facebook users has increased to a point where it has turned to be an addiction. The essential reasons behind the rising utilization of Facebook, which were found through investigating the mental thought processes, include improving sentiments of social interaction and self-preservation (Mukkamala et al., 2015).

Data is the basic form of information from which knowledge is formed after being mined and managed (Navathe & Ramez, 2000). Information is usually stored in its natural form known as text. The purpose of data mining is to completely locate trends and patterns from databases that were unfamiliar and

unknown (Gupta & Lehal, 2009). The techniques used in data mining include classification, clustering, neural networks, and decision trees (Sukanya & Biruntha, 2012). Data pre-processing is necessary prior to employing any technique. All the textual based information is electronically stored most probably in computers and web. The development of hardware storage devices has enabled storing and accessing large amount of data from laptops and computers. Websites give an intense medium for communication among people that prompt to common learning and sharing of valuable knowledge (Irfan et al., 2015; Ravi & Ravi, 2015). By joining various communities and discussion groups, people can communicate with each other on social networking websites such as Facebook, Twitter, and LinkedIn (Singh & Kumari, 2016). At whatever time and at any place, the issues that may emerge because of geographical separation can be solved by social networking. Unstructured or semi-structured language is widely used for communication on social networks. Ambiguities such as lexical, syntactic, and semantic errors arise when in regular day-to-day discussion individuals do not care about the spelling and grammatical structure and accuracy of a sentence and hence extracting logical patterns with precise information is required from such an unstructured form (Charalampakis et al., 2016). Huge data is accessible on the web in due to the recent swift expansion of social networks in utilization (Velasco-Elizondo et al., 2016). Person-to-person interaction behavior can uncover vital findings when text mining is applied on social networking sites data. Keeping in mind the end goal to locate a general opinion about a specific subject or human thinking patterns, the content mining technique can be utilized to investigate social networks.

## 2. Literature Review

Social media has rapidly changed the media landscape resulting in a competitive environment of news creation and dissemination (Mhamdi et al., 2018). Most of these news were provided in a textual format. Text mining has become one of the widespread fields in analyzing natural language processing (Salloum et al., 2018a). Sentiment analysis is regarded as an essential part of text mining. However, scholars have paid less attention to the sentiment analysis in the Arabic text (Salloum et al., 2018b). In this study, we considered analyzing the literature that is related to social media in general, and more specifically to the Arabic social media due to the limitation of published studies.

Content mining framework of social networking was suggested by Thomaz et al. (2016). Tests were carried out on this framework during the FIFA 2014 World Cup. Mining of Twitter contents which provided tourist with information on services such as hotels, conveyance and food were the main topic of interest of the study. Analysis of overall 58,686 appropriate messages was carried out and was gathered with application ontology. The analysis displayed an accurate real-time mirror image of the world tourism. This framework proved its efficiency in gathering relevant content and figuring out the famous topics on social networking sites regarding functional tourism administration. The study conducted a thorough review of the significance of sporting activities to learn about people's actions, and specifically about their sentiments. According to Gratch et al. (2015), samples from the 680 million tweets that were circulated throughout the 2014 World Cup has been collected and analyzed in order to identify the reasons why a sports activity is thrilling. Sentiment analysis method was employed and fascinating facts about sentiments attached to the people who want to see sports activity were obtained.

A workflow was designed by Chen et al. (2014) to incorporate both qualitative analysis as well as large-scale data mining methods. They concentrated on the tweets of engineering students to get an insight about the difficulties they go through in their university life. First, a qualitative analysis was conducted on the 25,000 tweets which were extracted in association with university life of the engineering students. Then, an algorithm was utilized to detect the issues faced by the students through 35,000 tweets made at the geo-location of Purdue University. Multi-label classification algorithms that classify tweets displayed the issues faced by the students. The major issues faced by these students were discovered; they include study burden, less social life and interactions and reduced hours of sleep.

In order to exchange current news, talking about complicated problems and promoting latest goods, China uses Sina Weibo which is quite popular over there. People get to know "what", "when", and "who" regarding popular subjects which are mostly tweeted as well as searched by several people by using Sina Weibo. Yu et al. (2014) examined hot topics Sina Weibo tweet data streams aiming at classifying sequential division of the hot topics which are browsed by the users of Sina Weiboduring. The study focused on a timeframe of four months to expose the interrelated hot topics which are tweets of same users and are also present in similar group of tweet messages. Moreover, to gain an insight of hot topics and the manner in which the users behave in various platforms, correlations among searches made on hot topic of the social media as well as on the search engines were analyzed. Hadoop MapReduce framework was discovered by Yu et al. (2014) in order to efficiently handle the massive amount of tweets obtained

from the gathered datasets, and measure the advantages of MapReduce to examine the tweet streams. Moreover, the difficulties that arise when examining huge tweet data were kept in mind. Certainly, this study is the first initiative to classify the sequential search patterns regarding hot topics on the platform of Sina Weibo.

As a result of extensive programs mining, topical experts on social media have become the center of attraction. A research was suggested and has incorporated information from four main social networking sites, they are: Twitter, Facebook, Google and LinkedIn, in addition to metadata, Wikipedia graph and Internet webpage text by (Spasojevic et al., 2016). A thorough examination of 37 features obtained from different mediums like Wikipedia, WebPages, message text, social graphs along with user lists was conducted. The research incorporated 12 billion messages or greater sliding window during a timeframe of 90 days as well as a 58 billion social graph edges. Options in Twitter Lists, Wikipedia, websites and Twitter Followers showed that these all are better constraints of proficiency. Around 90,000 labels were involved in the training of expertise ranking model by making use of these options on a huge ground-truth dataset. Application of this model is made on a production system which has ranking of around 650 million experts in at least more than 9000 topical areas every day. They provided outcomes and instances regarding the efficiency of the expert ranking system and experimental confirmation. At last, they formed a topical expertise data presented through an open REST APIs for extensive usage.

Blogs and social networks have recently become a valuable resource for studying sentiments in fields as diverse as customer relationship management, public opinion tracking and text filtering. As a matter of fact, knowledge obtained from social networks such as Twitter and Facebook has been shown to be extremely valuable to marketing research companies, public opinion organizations, and other text mining entity. However, Web texts have been classified as noisy as they represent considerable problems both at the lexical and the syntactic levels. Mostafa (2013) used a random sample of 3516 tweets to evaluate consumers' sentiment towards well-known brands such as Nokia, T-Mobile, IBM, KLM and DHL. He used an expert-predefined lexicon including around 6800 seed adjectives with known orientation to conduct the analysis. The results indicated a generally positive consumer sentiment towards several famous brands through using both a qualitative and quantitative methodology to analyze brands' tweets.

Sentiment Analysis as well as Text Mining offer the possibility of prediction and hotspot discovery on the Internet (Li & Wu, 2010). For the purpose of automatically analyzing the emotional polarity of a document, unsupervised Text Mining is required which can be accomplished by joining this algorithm along with K-means clustering and support vector machine (SVM). The suggested advance of the Text Mining was used to convert the group of data into clusters, where all centroids display a hotspot point in a present duration. Datasets having a variety of different topic points and 220,053 posts were obtained and edited from Sina sports blogs in order to utilize them in empirical research. It was observed from the outcomes that when SVM forecasting is used with K-means clustering algorithm, very reliable outcomes are obtained. 80% of K-means clustering outcomes were similar to 10 hotspot forums as mentioned by SVM prediction, where the outcomes for the top four hotspot forums were similar for SVM and K-means algorithm.

How text mining works for sentiment classification was studied by (Hamouda & Akaichi, 2013). A technique based on the Support Vector Machine (SVM) and Naïve Bayes algorithms was suggested for the objective of extracting helpful data regarding users' sentiments and way of reacting in critical situations (Dhanalakshmi et al., 2016). In the so-called "Arabic Spring" period Facebook posts of the Tunisian people were investigated, and the authors developed a sentiment lexicon that works according to the sentiments, short forms, and interjections from the statues that were obtained (Hamouda & Akaichi, 2013). Additionally, the authors carried out tests to make a comparison between SVM and Naïve Bayes machine learning algorithms by using training model for sentiment classification (Akaichi et al., 2013).

English is the main language used to conduct studies of sentiment mining. There are very few studies that tackle sentiment mining in Arabic language. Bilal et al. (2016) presented three classification models to carry out text classification by making use of the software known as Waikato Environment for Knowledge Analysis (WEKA). Points of views in Roman-Urdu and English language were obtained from a blog. Then, they were converted into the digital text format in order to form the training dataset that has 150 positive and 150 negative opinions, considered as instances. The outcomes clearly displayed that when comparisons are made among Decision Tree, KNN and Naïve Bayesian, Naïve Bayesian was the best in terms of F-measure, exactness, and accurateness, and recall.

Sobaci (2016) studied perspectives, point of views and sentiments from the casual chats of students on the social media platforms such as Twitter and Facebook. Information from these platforms was useful to

obtain knowledge about students. Though examining this information can be quite difficult, individual's understanding is needed to get through the complications present in the information in the social media. The increasing amount of information needs automatic data analysis methods.

Dos Santos and Ladeira (2014) described Usage of opinion mining on a dataset extracted from the Internet which contained the analysis of various Internet slangs, abbreviations and typing mistakes. Opinion mining is an area of research that seeks to figure out and classify the subjectivity, like point of views, perspectives or sentiments in natural language. Around 759,176 Portuguese reviews were extracted from the App Store of Google Play. Due to the huge quantity of reviews Frameworks such as Hadoop and Mahout from large-scale processing methods were utilized. A collection of 759 thousand reviews and a dictionary of slangs and abbreviations that are most often used on the web were made available as a result of this study.

Opinions are extracted from the textual data by using Sentiment analysis which also categorizes them into positive, negative or neutral classes. Impact caused on Arabic texts (tweets) by applying stemming as well as n-gram methods was studied by Brahimi et al. (2016). As a result, three classifiers, Support Vector Machines (SVM), Naive Bayes, (NB), and K-nearest neighbor (KNN), were used (Gelbukh, 2011). The findings revealed that the most excellent outcomes of performance were received during the application of hybrid demonstration which has tokens with character 3-grams. The outcomes of the test showed that accuracy of the three classifiers used in the working of opinion classification is considerably enhanced by making use of the feature selection method. Among all classifiers, SVM has showed the best performance. In addition, the top outcomes were obtained by using SVM as well as NB when the best options are selected from the SVM feature selection method.

Arabic language, a Semitic language, is known to be verbal form of communication in total in 57 countries by at least 422 million individuals as their first language and by 250 million as their second language (Al-diabat, 2012; Abdallah et al., 2012; Goudjil et al., 2013; Salloum et al., 2016; Ray & Shaalan, 2016). Arabic language ranks fifth among the most extensively used languages around the globe. Arabic language has a different grammatical structure according to morpho-syntactic analysis of the data which requires special focus for the purpose of providing the required data and reaching feasible outcomes (Alghamdi et al., 2014). Diacritics are known as particular symbols which are placed on the top or under the characters as per the knowledge of the sense and to assist in finding out the accurate pronunciation. Furthermore, they are used to differentiate among them, however they are eliminated initially by some Arabic morphological software, and they are used only when there are risks of complicated ambiguity in the documents that are intended for academic use. Stemming and light stemming are the two distinct methods of Arabic Language, where words are cut down to their stems in the process of stemming, where as in light stemming general affixes are eliminated from the words instead of cutting them to their stems. For example, in stemming the three Arabic words (المكتبة الكاتب الكتاب) which stand for (the library), (the writer), and (the book) in English language are reduced to one stem (كتب), that is (write). The aspect that several versions have different meanings or semantics despite the fact that they belong to the same root is the main reason to use light stemming (Duwairi et al., 2009).

Sentiment Analysis of contents in Arabic language is a relatively new area of research that is gradually attracting the attention of researchers. Due to the increasing availability of Arabic textual information through the web, it is urgently needed to process the relevant information using technological tools (Shaalán, 2014). Vast amount of knowledge is obtained from social media which is utilized by the government, organizations, and businesses to get understanding about people's perceptions. A semantic approach to discover people's perspectives and professional understandings from the medium of social media in Arabic language in both standard and dialects was suggested by Tartir and Abdul-Nabi (2016). The first edition of Arabic Sentiment Ontology (ASO) was launched by them and it comprises several words that express feelings and the manner such terms reveal the feelings. Furthermore, they demonstrated the usage of their approach in classifying various Twitter feeds on various topics.

Although substantial research on social media text mining has been conducted to extract relevant information from numerous sectors and changing it into useful knowledge, a new area of interest within text mining seems to be a good soil for research. Addressing textual analysis of Arabic newspapers' social networking websites can lead to interesting findings. Accordingly, this paper seeks to analyse online textual data and transform the quantifiable information into constructive knowledge through applying various text mining techniques in the context of newspapers' Facebook pages. Seeking to address this topic, this paper intends to answer the following research questions:

**Q1:** What are the common posts that were shared by people across the newspapers' Facebook pages?

**Q2:** Which country in the Arab Gulf region shares newspapers posts on Facebook most frequently?

**Q3:** Which newspaper shares posts on Facebook most frequently?

**Q4:** To what extent are the comments distributed in terms of their types?

**Q5:** What type of shared news is the most frequent in terms of shares count, comments count, and likes count?

### 3. Research Methodology

The methodology employed in this paper went through a process of data collection, data processing, and data analysis. 62327 posts from 24 newspapers pages on Facebook in the Arab gulf countries were extracted and analyzed. Posts with missing substance or inconsistent text have been removed when cleaning up the dataset during data processing. The Facepager software was used to collect the datasets. The query setup was utilized to collect the URLs. Subsequently, a local database would store the extracted data and it can be exported into a CSV format. 24 Arabic newspapers Facebook pages were short-listed to collect 62327 posts. To improve the data quality and performance, irrelevant attributes were excluded when importing the dataset into RapidMiner tool. To increase the precision, the missing attributes were removed from the dataset during the analysis stage. Some empty cells and special characters were observed in the collected data. According to the findings of **Atia** and Shaalan (2015), pre-processing techniques were employed to remove these special characters and empty cells.

### 4. Results

Building on the existing literature and applying various Text Mining techniques on the extracted data, the following findings were reached. The findings are presented and discussed in this section based on the research questions that are targeted by this paper.

**Q1: What are the common posts that were shared by people across the newspapers' Facebook pages?**

Around 9372 comments were generated from the highest shared post that refers to "**Live: New Year fireworks at Burj Khalifa**". This post got more than 10 thousand likes, with thousands of shares and comments. Several Text Mining techniques were applied on these comments in order to analyse the sentiments of the Facebook users. According to (Figure 1), the most frequently used terms by all users was "**Allah** الله" followed by "**Emirates** امارات", "**Year** عام", "**Good** خير", "**Save** يحفظ", "**Blessed** مبارك", and "**Graces** نعم", "**Happy** سعيد", "**Peace** سلام", and "**live** عاشت", respectively. The results revealed that the highest post contains a list of positive opinion terms (or positive sentiment words). We have utilized the comments data to analyze the views of people involved in various events, such as, people watching *Burj Khalifa*, Dubai New Year fireworks show, their participation in the festivals, and other allied events in the UAE. To evaluate the festival tourism provided by the tourists, the findings offer remarkable interpretations for the policy makers who are keen to develop the tourism sector in the dynamic and heterogeneous market of the UAE.

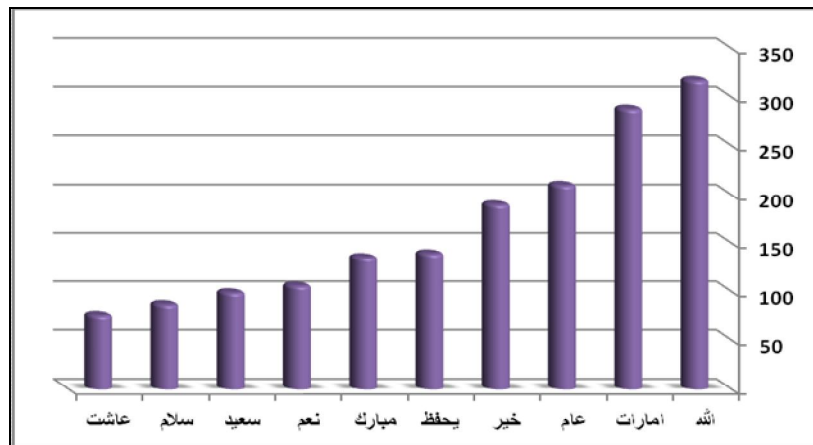


Figure 1. Comments count by Newspapers

**Q2: Which country shares Newspapers posts on Facebook most frequently?**

As per (Figure 2), it is noticed that UAE is the country that most frequently shares posts on Facebook, followed by Oman and KSA respectively. However, Qatar, Kuwait, and Bahrain were found to have the lowest share count of posts.

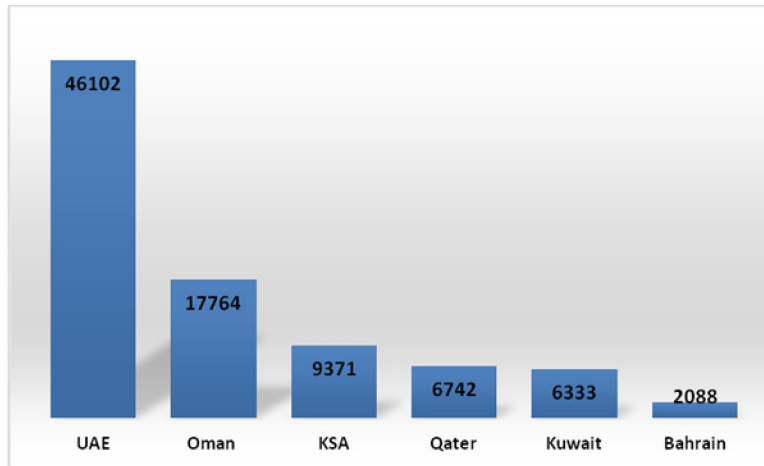


Figure 2. Share count by country

**Q3: Which newspaper shares posts on Facebook most frequently?**

According to the (Figure 3), it is noticed that *Albayan News* (UAE) is the newspaper that most frequently shares posts on Facebook, followed by *Alshabiba* (Oman), *Alkhaleej* (UAE) and *Emarat Alyoum* (UAE), respectively. In contrast, *Alwatanoman* (Oman), *Alqabas* (Kuwait), *Daralwatan* (Kuwait) and *Almustagbal* (Kuwait) were found to have the lowest share count of posts.

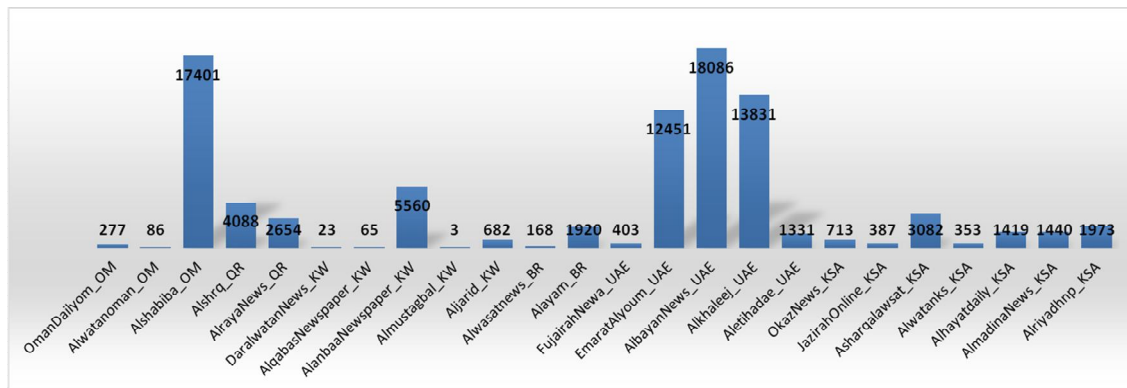


Figure 3. Share count by newspapers

**Q4: To what extent are the comments distributed in terms of their types?**

According to Zaza and Al-Emran (2015), a decision tree model has been implemented to predict different activities on Facebook. Figure 4 shows the terms nodes and leaves which have been mentioned in the research work of North (2012). The nodes are in the form of gray oval shapes. Such attributes work perfectly as predictors in our label attribute. Various categories of label attribute are categorized in the form of leaves with different colors that go from the branch of the tree to the end of the leaf. We can see in this tree that *comment count* is the best predictor of whether or not a person is going to comment on the newspapers' post. It is observed that video posts on Facebook attract people much more than other posts (photo, link, status).

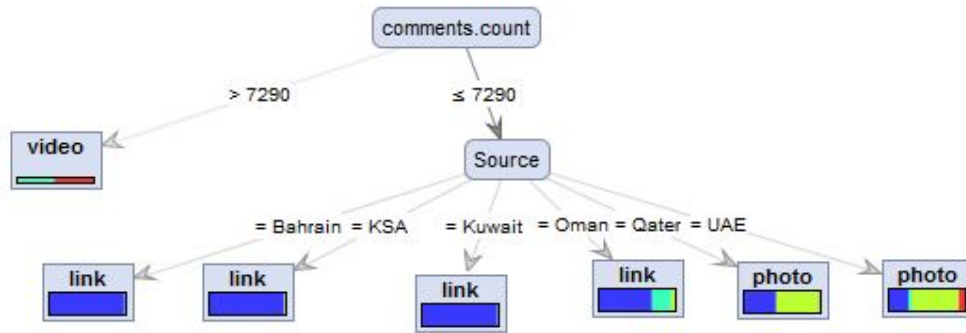


Figure 4. Decision tree of comments count

### Q5: What type of shared news is most frequent in terms of shares count, comments count, and likes count?

Facebook activities can be broken down into likes, comments, and shares, so that this information could be arranged for the interpretation of the data. Nonetheless, as a distinct marker, it seems that a good indication of content resonance is provided by the total Facebook interactions (the combined total of shares, likes, and comments). The illustrated figures depict how data mining can be employed to recognize the type of shared news in a data set that involves Facebook interactions (likes, comments and shares). (Figure 5.a) presents the *types* and their 'likes' count within the newspapers sources of data. The result shows that the *status* has the largest frequency rate as compared with other types presented in the group of newspapers. An updated feature enables users to discuss their opinions, share their locations or some information with friends is referred to as a Facebook status. Compared to the Twitter's based tweet on the social platform, a status is generally short and usually up to one or two lines. When a user updates his/her status, it is subsequently posted on his/her personal wall besides its reflection in the news feeds of their friends. A mobile site, a web browser or even a text message can be used to update the statuses. According to the (Figures 5.b & 5.c), *video* is the *type* that most frequently receives many comments and shares on Facebook. The question that arises here is why people are exchanging information through video sharing and why it is such a common and effective means to share the data? In due course, we share videos online so that others could respond to us. This response had different shapes. For instance, it can be in the form of comedy, gratitude, compassion, jolt, surprise and so on. Video sharing has become a rapid and a dominant way to show up and protest against social inequality wherever it is happening across the globe.

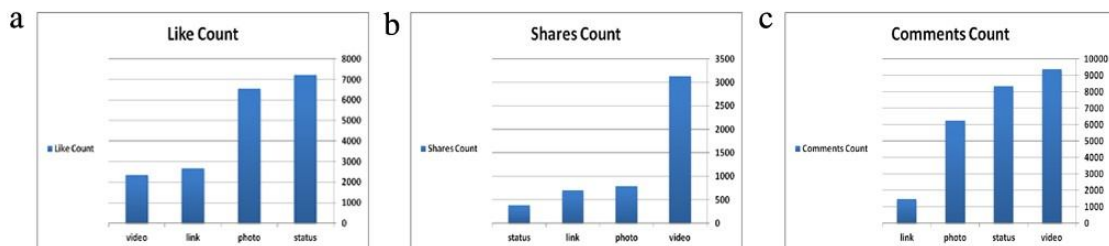


Figure 5. Likes, shares, and comments count by Newspapers

## 5. Conclusion

Text Mining has become one of the trendy research areas that was incorporated in different research fields such as Information Retrieval, computational linguistics, and Data Mining (Salloum et al., 2017a; Salloum et al., 2017b; Salloum et al., 2017c). Furthermore, text mining is an interrelated field with Natural Language Processing (NLP). NLP is one of the hot topics that is concerned with the interrelation among the huge amount of unstructured available text (Salloum et al., 2016; Salloum et al., 2017d), besides the analysis and interpretation of human-being languages (Al Emran & Shaalan, 2014; Al-Emran et al., 2015). Systems to naturally analyze Arabic computerized assets are not as effectively accessible as they are for analyzing English regardless of the expansion of Internet Arabic users.

This paper presents an endeavor to utilize data mining techniques on data extracted from Newspapers' Facebook pages including visualization, classification, clustering, word cloud and information retrieval. This study employed those techniques seeking to reach interesting knowledge and present it in different forms. 24 Arab Gulf newspapers' Facebook pages were investigated and analyzed through their 62327 posts. The findings revealed that the most commonly associated terms across all the newspapers are "Allah الله" followed by "Emirates امارات", "Year عام", "Good خير", "Save يحفظ", "Blessed مبارك", and "Graces نعم", "Happy سعيد", "Peace سلام", and "Iive عاشت". Additionally, the UAE is the country that most frequently shares posts on Facebook, followed by Oman and KSA while *Albayan News* (UAE) is the newspaper that most frequently shares posts on Facebook, followed by *Alshabiba* (Oman), *Alkhaleej* (UAE) and *Emarat Alyoum* (UAE). Arabic Text Mining and Arabic NLP will receive more research attention, we believe. A great deal of further research is still to be done. The researchers are interested in applying further text mining techniques on Arabic comments and posts to contribute to the existing literature of sentiment analysis.

## References

- Abdallah, S., Shaalan, K., & Shoaib, M. (2012, March). Integrating rule-based system with classification for arabic named entity recognition. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 311-322). Springer Berlin Heidelberg.
- Akaichi, J., Dhouioui, Z., & Pérez, M. J. L. H. (2013, October). Text mining facebook status updates for sentiment classification. In *System Theory, Control and Computing (ICSTCC), 2013 17th International Conference* (pp. 640-645). IEEE.
- Al Emran, M., & Shaalan, K. (2014, September). A Survey of Intelligent Language Tutoring Systems. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on* (pp. 393-399). IEEE.
- Al-diabat, M. (2012). Arabic text categorization using classification rule mining. *Applied Mathematical Sciences*, 6(81), 4033-4046.
- Al-Emran, M., Zaza, S., & Shaalan, K. (2015, May). Parsing modern standard Arabic using Treebank resources. In *Information and Communication Technology Research (ICTRC), 2015 International Conference on* (pp. 80-83). IEEE.
- Alghamdi, H. M., Selamat, A., & Karim, N. S. A. (2014). Arabic web pages clustering and annotation using semantic class features. *Journal of King Saud University-Computer and Information Sciences*, 26(4), 388-397.
- Al-Qaysi, N., & Al-Emran, M. (2017). Code-switching Usage in Social Media: A Case Study from Oman. *International Journal of Information Technology and Language Studies*, 1(1), 25-38.
- Atia, S., & Shaalan, K. (2015, April). Increasing the Accuracy of Opinion Mining in Arabic. In *2015 First International Conference on Arabic Computational Linguistics (ACLing)* (pp. 106-113). IEEE.
- Bilal, M., Israr, H., Shahid, M., & Khan, A. (2016). Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *Journal of King Saud University-Computer and Information Sciences*, 28(3), 330-344.
- Brahimi, B., Touahria, M., & Tari, A. (2016). Data and Text Mining Techniques for Classifying Arabic Tweet Polarity. *Journal of Digital Information Management*, 14(1), 15.
- Charalampakis, B., Spathis, D., Kouslis, E., & Kermanidis, K. (2016). A comparison between semi-supervised and supervised text mining techniques on detecting irony in greek political tweets. *Engineering Applications of Artificial Intelligence*, 51, 50-57.
- Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies*, 7(3), 246-259.
- Dhanalakshmi, V., Bino, D., & Saravanan, A. M. (2016, March). Opinion mining from student feedback data using supervised learning algorithms. In *Big Data and Smart City (ICBDSC), 2016 3rd MEC International Conference on* (pp. 1-5). IEEE.
- Dos Santos, F. L., & Ladeira, M. (2014, October). The role of text pre-processing in opinion mining on a social media language dataset. In *Intelligent Systems (BRACIS), 2014 Brazilian Conference on* (pp. 50-54). IEEE.
- Duwairi, R., AlRefai, M. N., & Khasawneh, N. (2009). Feature reduction techniques for Arabic text categorization. *Journal of the American society for information science and technology*, 60(11), 2347-2352.
- Gelbukh, A. (2011). *Computational linguistics and intelligent text processing*. Springer.
- Goudjil, M., Koudil, M., Hammami, N., Bedda, M., & Alruily, M. (2013, June). Arabic text categorization using SVM active learning technique: An overview. In *Computer and Information Technology (WCCIT), 2013 World Congress on* (pp. 1-2). IEEE.
- Gratch, J., Lucas, G., Malandrakis, N., Szablowski, E., Fessler, E., & Nichols, J. (2015, September). GOAALLL!: Using sentiment in the World Cup to explore theories of emotion. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on* (pp. 898-903). IEEE.



- Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1), 60-76.
- Hamouda, S. B., & Akaichi, J. (2013). Social networks' text mining for sentiment classification: The case of Facebook' statuses updates in the 'Arabic Spring' era. *International Journal of Application or Innovation in Engineering & Management (IJAEM)*, 2(5), 470-478.
- Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., & Tziritas, N. (2015). A survey on text mining in social networks. *The Knowledge Engineering Review*, 30(02), 157-170.
- Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision support systems*, 48(2), 354-368.
- Mhamdi, C. (2016). Transgressing media boundaries: News creation and dissemination in a globalized world. *Mediterranean Journal of Social Sciences*, 7(5), 272-277.
- Mhamdi, C. (2017). Framing "the Other" in Times of Conflicts: CNN's Coverage of the 2003 Iraq War. *Mediterranean Journal of Social Sciences*, 8(2), 147-153.
- Mhamdi, C., Al-Emran, M., & Salloum, S. A. (2018). Text Mining and Analytics: A Case Study from News Channels Posts on Facebook. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 399-415). Springer, Cham.
- Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.
- Mukkamala, R. R., Sørensen, J. I., Hussain, A., & Vatrappu, R. (2015, September). Social set analysis of corporate social media crises on facebook. In *Enterprise Distributed Object Computing Conference (EDOC), 2015 IEEE 19th International* (pp. 112-121). IEEE.
- Navathe, S. B., & Ramez, E. (2000). Data warehousing and data mining. *Fundamentals of Database Systems*, 841-872.
- North, M. (2012). *Data mining for the masses* (pp. 91-100). Global Text Project.
- Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- Ray, S. K., & Shaalan, K. (2016). A Review and Future Perspectives of Arabic Question Answering Systems. *IEEE Transactions on Knowledge and Data Engineering*, 28(12), 3169-3190.
- Salloum, S. A., Al-Emran, M., & Shaalan, K. (2016). A Survey of Lexical Functional Grammar in the Arabic Context. *International Journal of Computing and Network Technology*, 4(3).
- Salloum, S. A., Al-Emran, M., & Shaalan, K. (2017a). Mining Social Media Text: Extracting Knowledge from Facebook. *International Journal of Computing and Digital Systems*, 6(2), 73-81.
- Salloum, S. A., Al-Emran, M., & Shaalan, K. (2017c). Mining Text in News Channels: A Case Study from Facebook. *International Journal of Information Technology and Language Studies*, 1(1), 1-9.
- Salloum, S. A., Al-Emran, M., Abdallah, S., & Shaalan, K. (2017d, September). Analyzing the Arab Gulf Newspapers Using Text Mining Techniques. In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 396-405). Springer, Cham.
- Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017b). A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives. *Advances in Science, Technology and Engineering Systems Journal*.
- Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2018a). Using Text Mining Techniques for Extracting Information from Research Articles. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 373-397). Springer, Cham.
- Salloum, S. A., AlHamad, A. Q., Al-Emran, M., & Shaalan, K. (2018b). A Survey of Arabic Text Mining. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 417-431). Springer, Cham.
- Shaalán, K. (2014). A survey of Arabic named entity recognition and classification. *Computational Linguistics*, 40(2), 469-510.
- Singh, T., & Kumari, M. (2016). Role of Text Pre-processing in Twitter Sentiment Analysis. *Procedia Computer Science*, 89, 549-554.
- Sobaci, M. Z. (2016). *Social Media and Local Governments*. Springer International Publishing: Imprint: Springer.
- Spasojevic, N., Bhattacharyya, P., & Rao, A. (2016). Mining half a billion topical experts across multiple social networks. *Social Network Analysis and Mining*, 6(1), 63.
- Sukanya, M., & Biruntha, S. (2012, August). Techniques on text mining. In *Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on* (pp. 269-271). IEEE.
- Tartir, S., & Abdul-Nabi, I. (2016). Semantic Sentiment Analysis in Arabic Social Media. *Journal of King Saud University-Computer and Information Sciences*.
- Thomas, E. H., & Galambos, N. (2004). What satisfies students? Mining student-opinion data with regression and decision tree analysis. *Research in Higher Education*, 45(3), 251-269.
- Velasco-Elizondo, P., Marín-Piña, R., Vazquez-Reyes, S., Mora-Soto, A., & Mejia, J. (2016). Knowledge representation and information extraction for analysing architectural patterns. *Science of Computer Programming*, 121, 176-189.
- Yu, Q., Weng, W., Zhang, K., Lei, K., & Xu, K. (2014, December). Hot topic analysis and content mining in social media. In *Performance Computing and Communications Conference (IPCCC), 2014 IEEE International* (pp. 1-8). IEEE.

Zaza, S., & Al-Emran, M. (2015, October). Mining and Exploration of Credit Cards Data in UAE. In *e-Learning (econf), 2015 Fifth International Conference on* (pp. 275-279). IEEE.